BioE/MCB/PMB C146/246, Spring 2003

Problem Set 3: General Gap Penalties, Advanced Dynamic Programming, Substitution Matrices

Due 10 Feb 03, 5:00 pm PST by email to derek@rana.lbl.gov

1.   15 points
Sequence Alignment with Affine Gaps

Revise your alignment program from the last problem set using the following parameters:
    Identity   +4
    Transition   −2
    Transversion −4
    Gap   −8 (first position), −1 (each gap position)

Given the sequences

    ATTTTAAGCGCATACCGC
    TCGCAAATATAC

Perform a global alignment on the two sequences and report their score. Attach all dynamic programming matrices used (without tracebacks) to your email as *sid*_ps3_1.txt

2.   20 points
Perform any two of the following alignments, using the scoring matrix from Problem Set 2. You may use a program to assist you, though implementations are not required. Use gap penalties of −5 at each position, except where otherwise specified.
    (i)   Repeated matches, threshold 10   HECYDWH   and   HEWGH
    (ii)   Hirschberg/Myers/Miller alignment HECYDWH   and   HEWGH
    (iii)   Sub-optimal global alignment   HEAGAWGHE   and   PAWHEA
    (iv)   Global alignment with three-parameter gaps (−8, −1, −2) for the sequences
                SSFTLT   and   SCHKDIL

Include dynamic programming matrices, traceback paths, alignments and scores in your answer. Attach the dynamic programming matrix (without tracebacks) to your email as *sid*_ps3_2.txt

3.   5 points
Why are sub-additive gap penalties used? Give (at least) two reasons.

4.   5 points
Compare and contrast the construction and features of the BLOSUM and PAM series of matrices. Mention the strengths and weaknesses of each.

5.   5 points
A 1-PAM matrix changes on average of 1% of amino acids. Does a 2-PAM matrix change on average 2%? Explain.

6.   5 points
For alignments performed with PAM matrices, explain the meaning of a substitution score and the score of the alignment.

7.   5 points
Why are gap parameters NOT estimated the same way as substitution matrix parameters?

8.   15 points
Given the following BLOCK (multiple sequence alignment of proteins):

```
MMKE
MKKE
IKIE
MEME
IMKI
IKKE
MKME
IKKE
MKME
IKKE
```

(A) (10 points) Compute the joint probabilities q_{ij} and the marginal probabilities p_i for each i, j in the amino acid alphabet.

(B) (5 points) Compute the BLOSUM matrix for this BLOCK.

9.   15 points
Given the initial mutability matrix below, calculate the corresponding 3-PAM matrix. Normalize your answer such that each row and column sums to 1000.

	S	T	V
S	990	7	3
T	7	993	0
V	3	0	997